DOCUMENT RESUME

ED 130 642 IR 004 174

AUTHOR	Johnson, Kirk A.
TITLE	The Effect of Test Trials on the Learning of Paired Adjectives.
INSTITUTION	Aerospace Medical Research Labs, Wright-Patterson AFB, Ohio.
REPORT NO	AMRL-TR-64-105; P-1710; T-171007
PUB DATE	Oct 64
CONTRACT	AF-33(615)-1046; AF-33(657)-10234
NOTE	20p.; Prepared in part with the University of Cincinnati; Archival document
EDRS PRICE	MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS	College Students; Educational Research; Higher Education; Intermode Differences; *Paired Associate Learning; Prompting; *Teaching Machines

ABSTRACT

Three experiments were performed to extend the research using the Subject Matter Trainer (SMT), an early teaching machine. Various training modes were compared. In all three experiments, the subject's task was to learn adjective pairs. Experiment I used the anticipation method. A test slide was shown first with an adjective singled out at the top of the frame and a list of 20 potential response terms at the bottom. An answer slide was then presented with the correct response term underlined. Experiments II and III used a prompting method. Only the answer slides were shown during the learning trials. Four conditions were used in each experiment. In the "3L" condition subjects received three learning trials. In the "5L" condition they received five learning trials. In the "3LT" condition the subjects had alternating series of three learning and two test trials. In the "3LF" condition subjects experienced alternating series of three learning trials and two filler trials consisting of color and cartoon images requiring a rating of color intensity as the subject's response. The experiments varied as to the intertrial time interval and the time separating learning and retention series. In the learning series of each experiment, the "3LT" condition led to better learning than either "3L" or "3LF." The differences were not uniformly significant nor did they match in magnitude the differences reported in an earlier SMT study. No significant differences were found for the retention series. (KB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

TT-20

DB

ED130642

AMRL-TR-64-105

THE EFFECT OF TEST TRIALS ON THE LEARNING OF PAIRED ADJECTIVES

KIRK A. JOHNSON, PhD

OCTOBER 1964

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BEHAVIORAL SCIENCES LABORATORY
AEROSPACE MEDICAL RESEARCH LABORATORIES
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO

IR004174

FOREWORD

This study was initiated by the Behavioral Sciences Laboratory of the Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio. The research covered herein represents a portion of the applied research program of the Technical Training Branch, Training Research Division. The research was conducted under Contracts AF 33(657)-10234 and AF 33(615)-1046 with the University of Cincinnati, in support of Project 1710, "Training, Personnel and Psychological Stress Aspects of Bioastronautics," Task 171007, "Automated Training and Programed Instruction." The experiments were conducted by the University of Cincinnati under the direction of Dr. R. J. Senter, the principle investigator, and Dr. W. N. Dember. The author was technical monitor. This report covers research performed between July 1963 and August 1964.

The author is grateful to Dr. Gordon A. Eckstrand, Project Scientist, and Dr. Ross L. Morgan, Task Scientist, of the Behavioral Sciences Laboratory, for their ideas which suggested this study and for their helpful review of the final manuscript.

This technical report has been reviewed and is approved.

WALTER F. GRETHER, PhD
Technical Director
Behavioral Sciences Laboratory

ABSTRACT

Three experiments were conducted to determine the extent to which blocks of test trials contribute to the learning of paired adjectives. The results were compared to those found in an earlier study of the relative effectiveness of various modes of the Subject Matter Trainer. The earlier study indicated that learning by means of the Quiz Mode, under conditions in which blocks of trials in the Quiz Mode were alternated with blocks of test trials, was considerably superior to learning by means of several other modes, none of which involved test trials. In the present experiments, it was found that the addition of blocks of test trials to blocks of learning trials led to improvements in learning. This was true whether learning was by means of prompting or anticipation trials. The improvements were comparable in magnitude to the trial-by-trial differences found between the Quiz Mode and two of the remaining modes used in the previous study. Apparently, much of the trial by trial advantage of the Quiz Mode in the earlier study was due to the use of test trials interspersed between learning trials. Other evidence, however, indicates that when subjects are allowed to pace themselves, the Quiz Mode will lead to a considerably greater amount of learning in a given time than any of the other modes of the Subject Matter Trainer.

SECTION I

INTRODUCTION

In a study employing the Subject-Matter Trainer (SMT), an early teaching machine, Irion and Briggs (ref 9) found that, for paired associate materials and a fixed study time, the Quiz Mode of presentation was best of the four modes under investigation. The authors suggested that atleast part of this superiority might be attributed to the fact that in the Quiz Mode a student could elicit an indication of the correct alternative (from a field of 20) merely by pushing the quiz button; as a result, no time was lost in searching for correct responses, and more trials could be completed within the fixed interval of study time. They also suggested that the amount learned on each trial might have been greater under the Quiz Mode and indicated several reasons why this may have been the case.

A recent investigation by Kristofferson, Modrick, and Morgan (ref 12) was designed to test the second of these alternatives by comparing the number of trials required to reach successive criteria of mastery under the four modes employed in the previous study. One of the major obstacles to this approach was that the Quiz Mode provided no indication of the student's mastery of the material. To avoid this difficulty, a trial in the Quiz Mode was redefined to include not only a cycle of 20 items presentation under the standard Quiz Mode, but also a second cycle of 20 stimulus items to which the subject responded without formal feedback. Under this procedure, subjects learning by means of the Quiz Mode were found to require roughly half as many trials to reach the various criteria of mastery as did subjects learning by means of the Practice or modified Quiz Modes[1], the next most effective modes.

Since complete mastery was achieved in roughly half the time required by subjects in either the Practice or Modified Quiz Modes, it appears that this trial-by-trial superiority was not purchased at the price of an increased time per trial. The extent to which these results support a trial-by-trial superiority for the original Quiz Mode (ie, the Quiz Mode without the test blocks), however, is still subject to question because of the possible contribution of the test trials.

Unfortunately, there is little unanimity as to the probable effect of such tests. Estes (ref 4, p 211) states,

"It seems intuitively clear and can be demonstrated empirically (Estes,

---

[1]In the Practice Mode, the subject was required to guess at responses until he located the correct response. In the Modified Quiz Mode the subject was required to guess once at the correct response; he was then informed of the correct response just as he was in the Quiz Mode. Note that the Modified Quiz Mode, which was slightly inferior to the Practice Mode in the second SMT study (ref 12), was actually slightly superior to it in the first SMT study (ref 9). In neither case were the differences significant.

Hopkins, and Crothers, 1960), that no systematic increase in probability of correct responses to, say, paired-associate items will occur over a series of trials on which the stimulus members are presented alone and the S's responses receive no reward or informational feedback from the experimenter."

Similarly, Peterson and Brewer (ref 14, p 366) state:

"The proportion correct on a given presentation trial for the anticipation condition was approximately equal to the proportion correct on the preceding test trial, corroborating the finding of a short-term retention study (Peterson, Saltzman, Hillner, and Land, 1962) that the insertion of a test without knowledge of results produced no apparent change in the proportion correct on a later test when the outcome on the inserted test was ignored."

Most empirical evidence, however, favors some positive contribution from test trials, but the extent of this contribution varies widely from experiment to experiment.

The largest group of experiments bearing on this topic consists of studies in which subjects were tested repeatedly without feedback, following the completion of formal learning. When the terminal performance of such subjects is compared to the performance of control subjects who were tested but once after a comparable delay, the experimental subjects are usually found to have retained more material, sometimes to a significant extent (refs 5, 15), though more often at a level short of significance (refs 15, 16, 17). Other studies, however, have reported superior retention for the control group, again at levels both exceeding (ref 19) and failing (ref 16) the usual levels of significance.

Studies (refs 6, 7, 18, 22) in which significant gains in retention have been found between early and late tests in a given series provide somewhat more convincing evidence for the positive contribution of testing. Another study (ref 3) indicated that the latencies of correct responses decrease with repeated testing, but that the latencies of incorrect responses do not. Still another study (ref 10) demonstrated that following an initial error a correct response is more likely to be repeated than is a second error which differs from the first.

There is an obvious limitation to generalizations drawn from these studies to the S-T study: the studies considered above provide no indication of the effect which testing might have on subsequent learning[2]. This limitation is absent in the studies on percentage of occurrence of response members, or ORM.

---

[2]Estes (ref 4) recognized the possibility that test trials interspersed within a learning series might have an effect quite different from a like number of test trials at the end of the series, and offered evidence which indicated that such interspersed tests might be quite beneficial.

In these studies, subjects learn paired-associate material by the anticipation method. Variations in ORM are obtained by omitting varying percentages of the usual confirmations. Thus, 50% ORM would indicate a condition in which a given stimulus item, S_n, is followed by the correct response, R_n, on half of its occurrences and a blank frame on the remaining half. If the latter occurrences are not counted as trials[3], and if they have no effect on learning, then one would predict that 50% ORM would require the same number of trials to reach a given criterion of mastery as would 100% ORM. On the other hand, if the test trials contributed quite as much as the learning or confirmation trials, then one would predict that 50% ORM would require only half the trials that would be required under 100% ORM.

Over a number of variations in materials, list length, rate of presentation, and ORM, reductions in ORM have always led to a reduction in the number of confirmation trials required to reach mastery. In almost all cases, however, this reduction has been significantly less than that which would have been expected had the test trials contributed as much to learning as the confirmation trials. For studies in which nonsense syllables were used as response members, 50% ORM has been found to reduce confirmation trials to criterion by between 4% and 45%. The median savings, calculated over 13 comparisons (refs 6, 11, 21) was 29%.

For studies in which adjectives were used as response members, as was the case in the SMT study, 50% ORM has been found to reduce confirmation trials to criterion by between 26% and 46%. The median savings, calculated over 6 comparisons (including one study at 40% ORM and one at 60% ORM (refs 11, 20, 21)) was 43%. If one generalized from this figure to the SMT study, then one would predict a savings quite close to the 47%[4] difference between the Quiz Mode and the next most efficient mode.

The ORM studies involve a random mixing of the test and confirmation items. In the SMT study, however, the test items were grouped together into blocks. There are at least three studies which should shed light on the effect of such blocking procedures, but unfortunately there is little agreement among the three. Battig and Brackett (ref 1) required their subjects to learn a set of 12 nonsense shape-number pairs by means of either the standard anticipation

---

[3]To facilitate comparisons with the SMT studies, the results of the ORM studies have been transformed, when necessary, to the results which would have been obtained had only the confirmation trials been counted.

[4]The mean number of trials to criterion were actually 4.3 for the Quiz Mode and 8.1 for the Practice Mode. These may, however, provide a somewhat biased picture of acquisition. Test 1 for the Quiz Mode reflects the subjects' learning on the first trial in the Quiz Mode, whereas Trial 1 for all other modes probably reflects little more than the subjects' ability to guess. If one corrected for this bias by using performance on the second trial under the Practice Mode as an index of first trial learning, that is, by omitting the first trial in the Practice Mode, then the percent advantage of the Quiz Mode would drop to 39%.

method or the recall method. In the latter method, a trial consisted of a block of SR pairs presented simultaneously (prompt items) followed by a block of stimuli presented without confirmation (test items). Since time per presentation was held constant, trials under the two methods were or approximately the same duration.

Battig and Brackett found that subjects learning under the recall method averaged 35% fewer trials to reach criterion than subjects learning under the anticipation method. This savings is roughly the same as the savings found in the ORM studies, and indicates an effect which remains constant across variations, not only in the grouping of test trials, but also in duration of presentation (the test trials required additional time in the ORM studies), mode of presentation (anticipation plus test in the ORM studies, prompt plus test in this study), and materials.

In a second study, those same authors (ref 2) investigated the interactions between learning method and ORM. They found that in 25 trials under 100% ORM, subjects learning by means of the recall method made 11% more correct responses than subjects learning by means of the anticipation method. The use of correct responses instead of trials to criterion precludes exact comparisons between this study and most of the studies considered previously[5], but the direction of the difference is in keeping with the direction found in the first study by these authors.

For 25 trials under 50% ORM, however, they found that subjects learning by means of the anticipation method actually did better than those learning by means of the recall method. The authors explained this difference between 100% and 50% ORM by means of a hypothetical facilitatory effect due to the "complete temporal separation of presentation of S-R pairs for learning and S terms alone for recall" (ref 2, p 513). This effect was able to operate during the recall trials under 100% ORM, but was destroyed by the shift to 50% ORM.

The most deviant of the three studies bearing on the effect of blocking the test trials was reported by Lockheed (ref 13). Subjects were required to learn a list of 9 paired nonsense syllables by means of the anticipation method, the recall method, or the random method. The latter was similar to 50% ORM,

---

[5]Negatively accelerated learning curves of the type found in these studies will generally produce a larger percent difference in trials to criterion than in correct responses over a fixed number of trials. For example, in the SMT study, the advantage of the Quiz Mode over the Practice Mode in terms of a reduction in trials to criterion was 39% (the first trial on the Practice Mode being omitted on the grounds that it reflects nothing but guessing behavior). When total correct responses are compared (again omitting the first trial on the Practice Mode) after each of the first four trials, however, the advantages of the Quiz Mode are only 18%, 18%, 14%, and 14%, respectively.

except that prompting was used instead of anticipation, and all items, whether S alone or SR simultaneously, were of the same duration. He found that all subjects in the anticipation and random conditions reached a criterion of two perfect repetitions within the allotted 60 trials, but that 8 out of the 23 subjects in the recall group failed to do so. He also found that subjects in the anticipation group reached criterion somewhat more rapidly than those in the random group.

These last two studies cast some doubt on the contribution to learning of test trials administered in blocks. Even if one accepts the position that there is, in general, some positive contribution, one has no data that would permit one to estimate the magnitude of such a contribution in the case of meaningful materials such as were used in the second SMT study. Without such information, one cannot estimate the trial by trial advantage of the Quiz Mode, or for that matter, assert with confidence that such an advantage exists.

The following studies were designed to provide information on the effect that blocks of test trials have on the learning of paired adjectives. In all cases, the procedures of the SMT study have been modified to permit the testing of subjects in groups.

SECTION II

METHOD

PROCEDURE

Three experiments were performed. In all 3, the subject's task was to learn 20 adjective pairs of low intra-pair rated similarity (ref 8). The pairs of adjectives were the same as those used in the second SMT study (ref 12). All materials were presented by means of 35 mm slides. Intratrial timing was achieved by slaving a Kodak Carousel projector to a Graflex Audio-Graphic Instructor.

Two types of slides were used to present the adjectives. A Test Slide consisted of a single stimulus term centered in the upper third of the slide and 20 alternative response terms arranged in 4 columns of 5 terms each. The latter occupied the lower two-thirds of the slide. An Answer Slide differed from a Test Slide in that the response term which was correct for the stimulus in question was bordered by a black frame. The arrangement of response terms was the same for all slides of either type.

Two additional types of slides were used to present filler materials. A Color Slide consisted of a single colored rectangle. A Cartoon Slide consisted of a cartoon reproduced from a magazine.

In experiment I, the test and answer slides were arranged so that the subjects learned by the anticipation method. That is, a Test Slide was shown first and the subject tried to covertly select the correct response term. Following the Test Slide, the appropriate Answer Slide was shown indicating the correct

response item. In experiments II and III, learning was by the prompt method, ie, only Answer Slides were shown during learning trials. The time intervals for the various Slides in the three experiments are shown in figure 1.

In all experiments, a test item involved the presentation of a Test Slide and required the subject to write his choice as to the correct response. The only other overt responses required of the subjects were their rating of the "intensity" of the colors used as filler items in experiment I. Each color was rated while it was presented.

All 20 adjectives were presented on each of the learning and test trials. The order of presentation remained constant over all trials. The actual times required for the various conditions in each of the three experiments can be found in figure 2.

Four conditions were used in each of the three experiments. In the 3L condition subjects simply received three learning trials. In the 5L condition they received five learning trials. In the 3LT condition the subjects received an alternating series of three learning trials and two test trials. In the 3LF condition the subjects received an alternating series of three learning trials and two filler trials.

The types of trials employed and the duration of each are indicated in figure 2. In experiment I, 1 week separated the learning and retention series. The duration of the prompted items of experiments II and III was considerably less than that of the covert anticipation items of experiment I; the duration of test items was also somewhat less in experiments II and III. In experiment II, only 1 test trial was used in the retention series and approximately 7 weeks separated the learning and retention series. In experiment III, the intertrial interval was reduced from approximately 100 sec. to approximately 50 sec. Two test trials were given at the completion of the learning series, and two test trials were given in the retention series. Three weeks separated the learning and retention series.

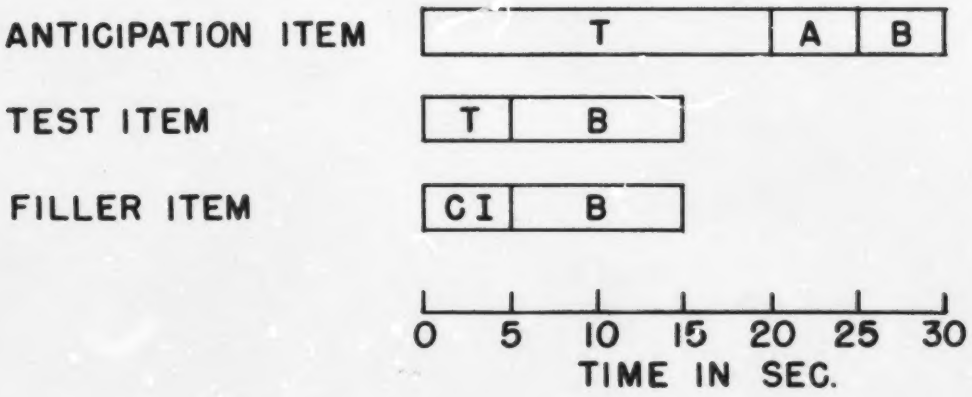
SUBJECTS

In experiments I and II all subjects were regular students from the University of Cincinnati. They were randomly assigned to the various experimental conditions and were run in groups of from approximately four to six per session. In experiment III, all subjects were from 2 Introductory Psychology classes in the University of Cincinnati Evening College. Each of the 2 classes was randomly split into 2 subgroups and these subgroups were randomly assigned to the 4 learning conditions. All members of a group were run at the same time. The number of subjects in each group for the various experiments is indicated at the base of each bar in figure 3.

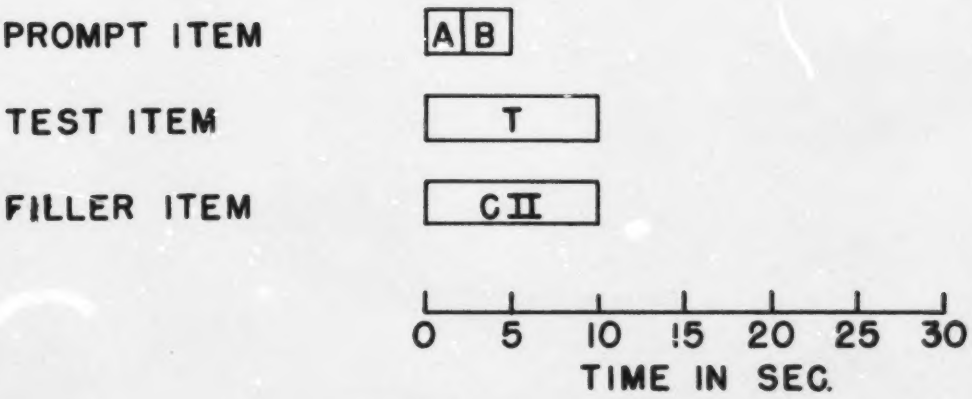
RESULTS

Figure 3 indicates the mean number of items correct on the final learning trial and the final retention trial for each group in each of the three experiments. Kruskal-Wallis tests were applied to each of these six trials. Significant differences were found among the learning series scores of both experiment II ($p < .01$) and experiment III ($p < .05$). For the two analyses on which significant

EXPERIMENT I



EXPERIMENTS II and III



KEY TO SLIDE TYPES

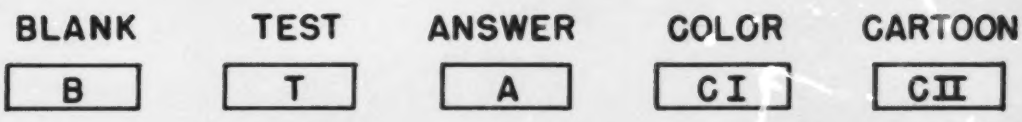


FIGURE 1. DURATION AND COMPOSITION OF ITEMS USED IN VARIOUS EXPERIMENTS

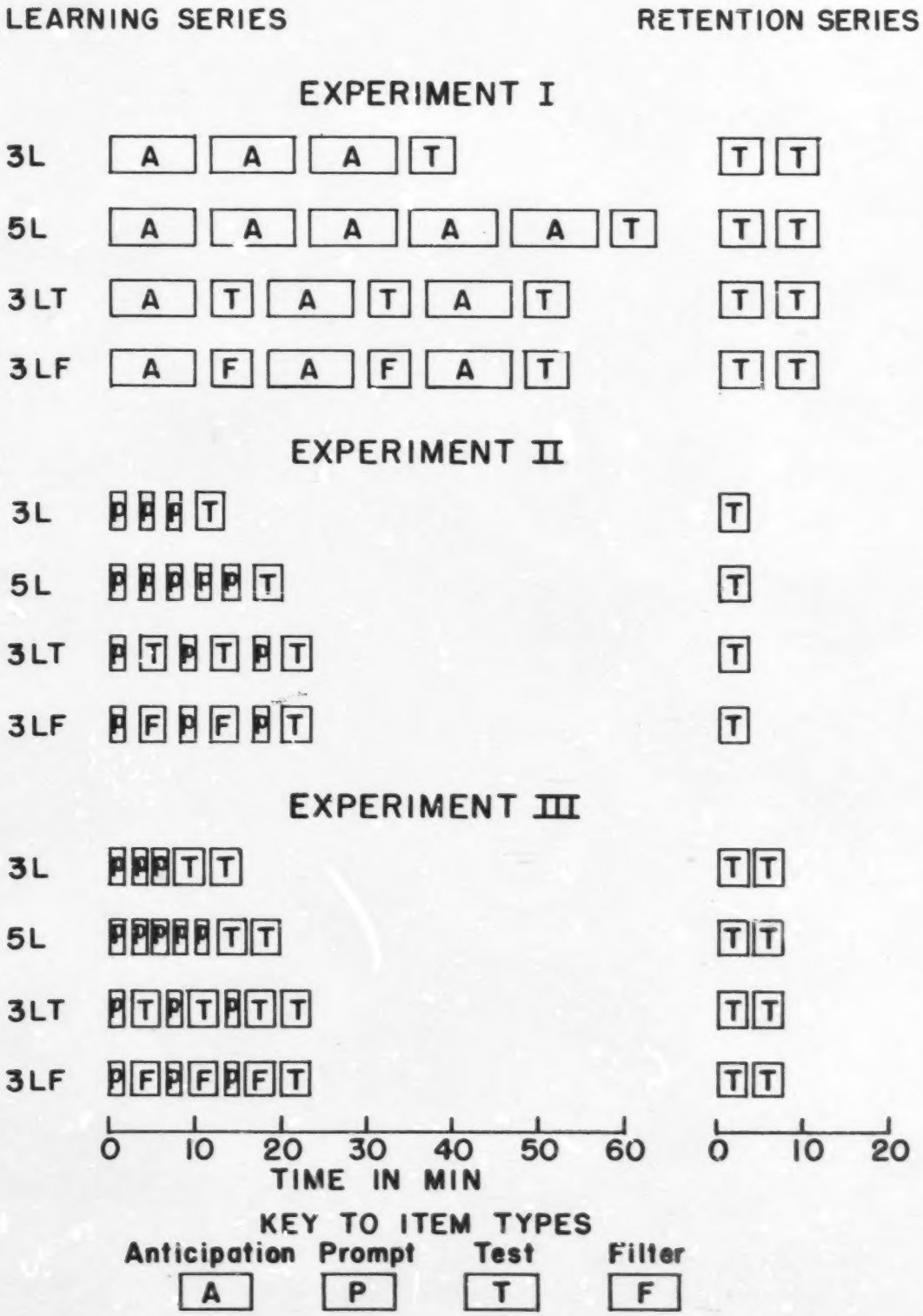
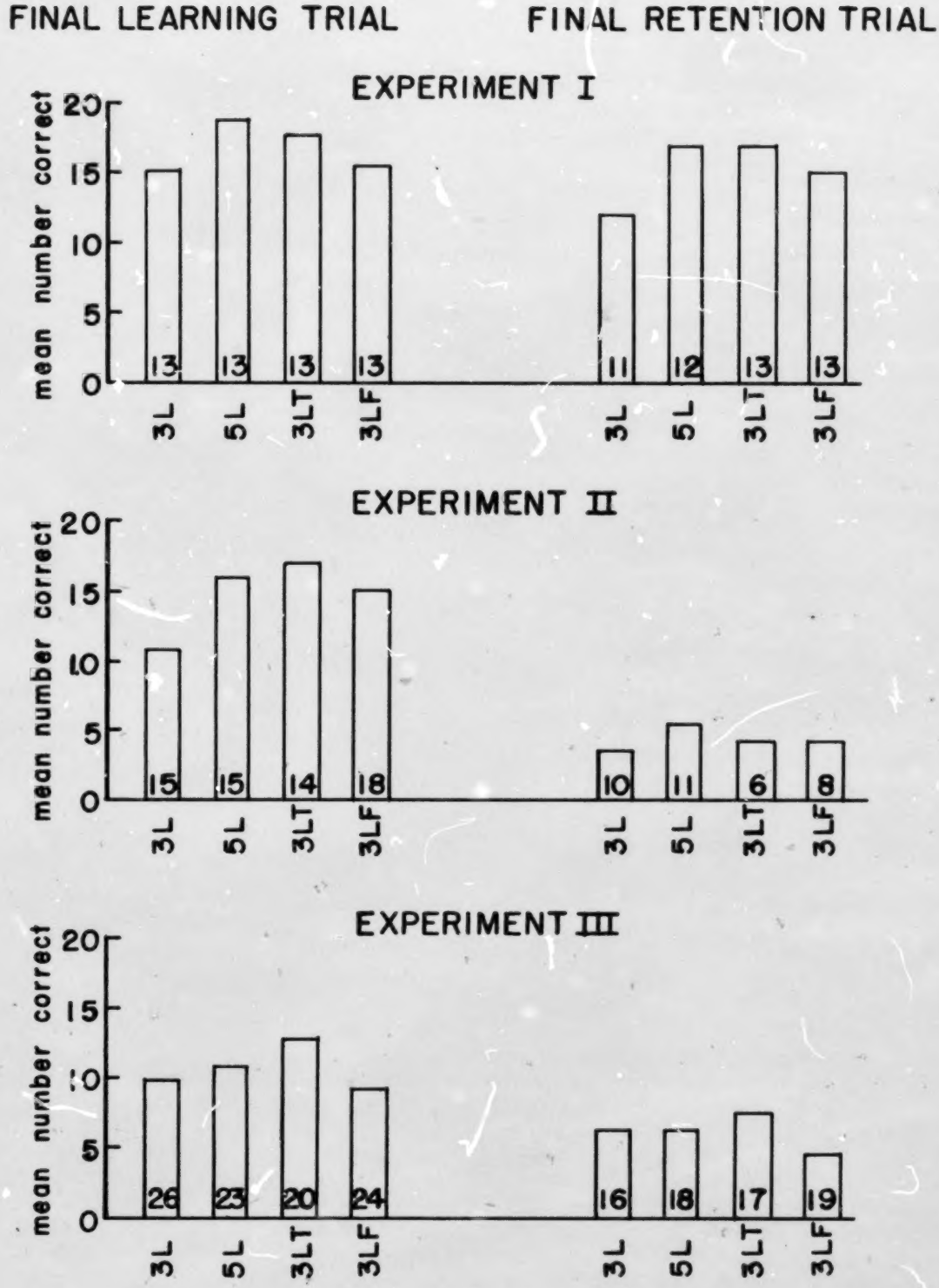


FIGURE 2. DURATION AND COMPOSITION OF TRIALS USED IN VARIOUS EXPERIMENTS



Note: Number in base of each bar indicates the number of subjects in that group.

FIGURE 3. MEAN NUMBER CORRECT ON FINAL LEARNING TRIAL AND FINAL RETENTION TRIAL

differences were found, Mann-Whitney U tests were applied to individual pairs of learning conditions. In the learning series of experiment II, the 3L group was significantly inferior to all other groups ($p < .01$ in all cases). In the learning series of experiment III, the 3LF group was found to be significantly inferior to the 3LT group ($p < .05$). No significant differences were found for the retention series.

In experiment I, subjects improved an average of 0.63 items from the first to the second trial in the retention series. In spite of the fact that 14 subjects made perfect scores on the first of these trials, and so could not improve, a Wilcoxon signed-ranks test, computed over the 4 learning conditions, indicated that the improvement was statistically significant ($p < .01$).

In experiment III, subjects again showed an improvement over the retention trials, this time one which averaged 1.30 items. A Wilcoxon signed-ranks test indicated that the improvement was again statistically significant ($p < .01$). A similar comparison of the last test trials in the learning series of this same experiment indicated that the subjects actually dropped an average of 0.34 items from the first to the second. This loss, however, did not reach an acceptable level of statistical significance ($0.10 < p <$

SECTION III

DISCUSSION

The differences found did not reach statistical significance as often as originally anticipated (even the difference between 3L and 5L was significant in but one comparison). Nevertheless, it is unlikely that further experimentation would modify to any appreciable extent the general conclusions which one can draw from the experiments presented above.

Before proceeding to a more detailed comparison between the present findings and those of the second SMT study (ref 12) it is necessary to indicate two probable errors in the description of the time data in the report of the second SMT study. An examination of the raw data indicates that many of the subjects in the Quiz Mode were reported to have spent 0 time on one or more trials. Since the pure mechanics of stepping the SMT through a series of 20 items requires nearly a minute, these raw data suggest that (a) on trials likely to be criterion trials, certain subjects were permitted to omit the prompted trials of the Quiz Mode and instead, to proceed directly from one test block to the next, and (b) that the time required for at least some of the test blocks was not included in the total times reported. The time data in the earlier report (ref 12) do not consistently include the time for both learning trials and test trials as stated in the report. Apparently, the Quiz Mode's advantage in terms of total time is considerably less than that which was reported, and could conceivably be nonexistent. This possibility should be borne in mind when considering the time differences which exist in the present series of studies.

In the learning series of each of the present experiments, the 3LT condition led to better learning than did either the 3L or the 3LF condition. In the

first of the three experiments, neither of these differences was significant. At least part of this failure, however, can probably be attributed to the fact that in this experiment there was insufficient opportunity for variations at the higher levels of mastery to manifest themselves; 28 of the 52 subjects made perfect scores.

The differences found in the present studies were not uniformly significant, nor did they match in terms of magnitude the differences in trials to criterion reported in the second SMT study (ref 11). When the latter study is viewed in terms of the differences present after a given number of trials, however, it is found to be more comparable. Table I indicates the mean number of items correct on the final trial of the learning series for groups 3LT, 3LF, and 3LT in each of the present experiments. It also indicates the percent advantage provided by the test trials in 3LT when computed against the performance of groups 3LF and 3L. For comparison, table II indicates mean number of items correct on Trials 1, 2, and 3 for the groups learning under the Quiz, Practice, and Modified Quiz Modes of the second SMT study[6]. It also indicates the percent advantage provided by the test trials in the Quiz Mode, when computed against the performance of the groups learning under th Practice and Modified Quiz Modes. The most valid comparison for a given study in the present series would be with that trial of the second SMT study on which the subjects had reached approximately the same level of mastery. The exact determination of such a point is impossible, but the 3 trials reported would seem to cover the appropriate range. In general, the percent advantage scores obtained in the present series of experiments are quite similar to those obtained in the SMT study. The most likely candidate for a truly greater difference in the SMT study is the 72% advantage of the Quiz group over the Modified Quiz group on Trial 1. Note, however, that the Quiz group spent 75% more time on Trial 1 than did the Modified Quiz group.

Not only the magnitude, but also the significance of the differences found in the SMT study are reduced when performance is viewed on a trial by trial basis. The Quiz and Practice groups fail to differ significantly on the first and third trials; the difference is but marginally significant on the second trial ($p = .048$).

Although none of the overall analyses on the retention series produced an indication of significant differences, the general pattern of the results would once again seem to offer some support for the contention that blocks of test trials aid learning. In each of the experiments, the 3LT group was superior to the 3L group, and in two of the experiments, the 3LT group was superior to the 3LF group. For the single exception, Experiment II, the superiority of the 3LF group was quite small (.21 items), as was the number of subjects involved. Comparisons between individual treatment pairs, though somewhat irregular in the absence of overall indications of significance, lend some additional support to this belief. Mann-Whitney U tests indicated that the 3LT

---

[6]The data for Trial 1 of the Practice and Modified Quiz Modes were taken from the subjects' performance on Trial 2, since, as was indicated in footnote 4, these data reflect what the subjects learned on Trial 1. Similarly, the data for Trial 2 were taken from the subjects' performance on Trial 3, and those for Trial 3 were taken from Trial 4.

TABLE I

MEAN NUMBER CORRECT FOR THREE METHODS OF PRESENTATION AND PERCENT DIFFERENCES

	Experiments		
	I	II	III
3LT	17.77	16.93	12.70
3LF	15.46	14.89	9.08
3L	15.15	10.80	9.81
3LT-3LF/3LF	.15	.14	.40
3LT-3L/3L	.17	.57	.31

TABLE II

MEAN NUMBER CORRECT FOR THREE MODES OF PRESENTATION AND PERCENT DIFFERENCES FROM THREE TRIALS OF SMT EXPERIMENT

	Trial 1	Trial 2	Trial 3
Quiz (Q)	13.42	17.08	18.42
Practice (P)	11.42	14.42	16.13
Modified Quiz (MQ)	7.79	11.29	13.58
Q-P/P	.18	.18	.14
Q-MQ/MQ	.72	.51	.36

group was superior to the 3L group in experiment I and to the 3LF group in experiment III ($p < .05$ in both cases).

The preceding discussion has dealt with the effect of test trials interspersed within the learning series. The present experiments also provide evidence bearing on the effect of repeated test trials administered after the completion of the learning series. The significant improvements found between the first and second test trials in the retention series of experiments I and III are what one might have expected from the similar improvements found between the initial trials of delayed test series in the experiments of Richardson (ref 18) and Wollen (ref 22). Richardson (ref 18) suggests that such an increase might be the result of the warmup provided by the early test trials.

Although it is almost certain that such warmup contributes to the present results, Goss, Morgan, and Golin (ref 6) found a similar increase over repeated test trials under circumstances which would seem to preclude an appeal to warmup. They shifted their subjects from the learning to the test series without any intervening delay during which subjects might be expected to lose their set. Also, most groups learned under something less than 100% ORM, and so had ample opportunity to become familiar with test items before entering the test series.

Even so, a significant improvement was found over repeated tests. In fact, several groups still appeared to be improving after as many as 16 test trials, and 6 of the 8 groups reached a level above that which they reached on their criterion trials[7]. Similar results (with the exception of the final superiority over criterion trials, which in this case were set at perfect performance) were found under similar circumstances in a second experiment by Goss, Nodine, Gregory, Taub, and Kennedy (ref 7).

The decrease in performance between the last two test trials in the learning series of experiment III, though not significant, does indicate a situation which differs from that found in the retention series and so represents something of a problem. A greater gain in the retention series might have been expected as the result of warmup, but the actual decline in performance, if reliable, would appear to be incompatible with the generally beneficial effects produced by test trials in both the learning and retention series of these and other experiments.

---

[7]It should be noted that two opposing factors would tend to bias this last comparison. On the one hand, some learning probably occurred on the last criterion trial and this learning would not have been reflected in performance on that trial. On the other hand, performance on the criterion trials was probably determined to a certain extent by lucky guesses, ie, a guess might cause a trial which would not otherwise be a criterion trial to become a criterion trial, thus raising the ratio of lucky to unlucky guesses on such trials to a level above chance.

Several of the experiments in which subjects were tested repeatedly immediately upon the completion of the learning trials (eg, refs 6, 7, 10), do not report results in a manner that would permit direct trial by trial comparisons of number correct. Two miniature experiments of the RTT[8] variety, do, however permit such comparisons: Jones (ref 10) reports a drop from 53 to 48 percent correct over the two test trials; Eimas and Zeaman (ref 3) report a drop from 57 to 55 percent.

Estes, Hopkins, and Crothers (ref 5) report another miniature experiment which provides four comparisons relevant to this point. With an RTRTT paradigm they found, in one case, an increase from 77 to 78 percent correct; in another, a drop from 83 to 78 percent. With an RRTT paradigm they found, in one case, a drop from 70 to 67 percent; in another, that subjects scored 73 percent correct on both test trials.

These findings would seem to offer some additional support to the belief that, in contrast to the improvement in performance which is commonly found over the first two trials of a test series initiated after some delay, one can expect a drop in performance over the first two trials of a test series initiated immediately upon the completion of learning.

SECTION IV

CONCLUSION

In the second SMT study (ref 12), it was found that there was a trial by trial superiority of the Quiz Mode over all additional modes. The effects of intrinsic differences among the modes, however, were confounded with the effects produced by the addition of test blocks between blocks of items in the Quiz Mode; none of the remaining modes required such test blocks.

In the present series of experiments, the addition of test blocks to blocks in which learning was by either prompting or anticipation led to improvements in performance which were comparable in magnitude to the differences found between the Quiz Mode and two of the remaining modes employed in the previous study.

It would seem more parsimonious to attribute much of the trial by trial superiority of the Quiz Mode, as found in the second SMT study (ref 12), to the effect of the test blocks rather than to intrinsic differences among the modes. This does not, however, reduce the practical importance of the Quiz Mode's greater efficiency (in terms of degree of mastery per unit time) demonstrated in the first SMT study (ref 9).

---

[8]An "R" trial is one on which the stimulus response pairs are presented simultaneously. A "T" trial is one on which only the stimulus items are presented. The subject responds, but receives no feedback from the experiment (ref 4).

LIST OF REFERENCES

1. Battig, W.F., and Brackett, H.R., "Comparison of Anticipation and Recall Methods in Paired-Associate Learning," Psychological Report 9: 59-65, 1961.

2. Battig, W.F., and Brackett, H.R., "Transfer from Verbal-Discrimination to Paired-Associate Learning: II. Effects of Instralist Similarity, Method, and Percentage Occurrence of Response Members," Journal of Experimental Psychology 65: 507-514, 1963.

3. Eimas, P.D., and Zeaman, D., "Response Speed Changes in an Estes' Paired-Associate 'Miniature' Experiment," Journal of Verbal Learning and Verbal Learning and Verbal Behavior 1: 384-388, 1963.

4. Estes, W.K., "Learning and the New 'Mental Chemistry'," Psychological Review 67: 207-223, 1960.

5. Estes, W.K., Hopkins, B.L., and Crothers, E.J., "All-or-None and Conservation Effects in the Learning and Retention of Paired Associates," Journal of Experimental Psychology 60: 329-339, 1960.

6. Goss, A.E., Morgan, C.H., and Golin, S.J., "Paired-Associates Learning as a Function of Percentage of Occurrence of Response Members (Reinforcement)," Journal of Experimental Psychology 57: 96-104, 1959.

7. Goss, A.E., Nodine, C.F., Gregory, B.N., Taub, H.A., and Kennedy, K.E., "Stimulus Characteristics and Percentage of Occurrence of Response Members in Paired-Associates Learning," Psychological Monographs 67, (12, Whole No. 531), 1962.

8. Hilgard, E.R., "Methods and Procedures in the Study of Learning," Handbook of Experimental Psychology, S.S. Stevens (Editor), pp 517-567, Wiley, New York, 1951.

9. Irion, A.L., and Briggs, L.J., Learning Task and Mode of Operation Variables in Use of the Subject-Matter Trainer, Technical Report AFPTRC-TR-57-8, Air Force Personnel and Training Research Center, Lackland AFB, Texas, 1957.

10. Jones, J.E., "All-or-None versus Incremental Learning," Psychological Review 69: 156-160,1962.

11. Klauser, D.H., McLaughlin, B.S., and Kulik, J.A., "S-R and R-S Paired-Associate Learning with Varying Occurrence of Response Members," Journal of Verbal Learning and Verbal Behavior 1: 85-90, 1962.

12. Kristofferson, A.B., Modrick, J.A., and Morgan, R.L., A Comparison of Modes of Presentation of Paired-Associates on the Subject-Matter Trainer, AMRL Memorandum P-46, 6570th Aerospace Medical Research Laboratories, Wright-Patterson AFB, Ohio, 1963.

13. Lockheed, G.R., "Methods of Presenting Paired Associates," Journal of Verbal Learning and Verbal Behavior 1: 62-65, 1962.

14. Peterson, L.R., and Brewer, C.L., "Confirmation, Correction, and Contiguity," Journal of Verbal Learning and Verbal Behavior 1: 365-371, 1963.

15. Peterson, L.R., Saltzman, D., Hillner, K., and Land, V., "Recency and Frequency in Paired-Associate Learning," Journal of Experimental Psychology 63: 396-403, 1962.

16. Postman, L., and Phillips, L.W., "Studies in Incidental Learning: IX. A Comparison of the Methods of Successive and Single Recalls," Journal of Experimental Psychology 61: 236-241, 1961.

17. Reynolds, J.H., "Repetition Factors in One-Trial Learning of Paired Associates," Journal of Experimental Psychology 65: 362-367, 1963.

18. Richardson, J., "The Relationship of Stimulus Similarity and Number of Responses," Journal of Experimental Psychology 56: 478-484, 1958.

19. Rohrer, J.H., "Factors Influencing the Occurrence of Reminiscence: Attempted Formal Rehearsal During the Interpolated Period," Journal of Experimental Psychology 39, 484-491, 1949.

20. Shulz, R.W., and Runquist, W.N., "Learning and Retention of Paired Adjectives as a Function of Percentage Occurrence of Response Members," Journal of Experimental Psychology 59: 409-413, 1960.

21. Wilcoxon, H.C., Wilson, W.R., and Wise, D.A., "Paired-Associate Learning as a Function of Percentage of Occurrence of Response Members and Other Factors," Journal of Experimental Psychology 61: 283-289, 1961.

22. Wollen, K.A., "One-Trial Versus Incremental Paired-Associate Learning," Journal of Verbal Learning and Verbal Behavior 1: 14-21, 1962.